SSOAR Open Access Repository
www.ssoar.info

Methodological Pitfalls of Convergence Analysis

Holzinger, Katharina

Postprint / Postprint
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:
Holzinger, K. (2006). Methodological Pitfalls of Convergence Analysis. *European Union Politics*, *7*(2), 271-287. https://doi.org/10.1177/1465116506063721

Nutzungsbedingungen:
Dieser Text wird unter dem "PEER Licence Agreement zur Verfügung" gestellt. Nähere Auskünfte zum PEER-Projekt finden Sie hier: http://www.peerproject.eu Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:
This document is made available under the "PEER Licence Agreement ". For more Information regarding the PEER-project see: http://www.peerproject.eu This document is solely intended for your personal, non-commercial use.All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.
By using this particular document, you accept the above-stated conditions of use.

gesis Leibniz-Institut für Sozialwissenschaften

Mitglied der Leibniz-Gemeinschaft

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-229022

European Union Politics
DOI: 10.1177/1465116506063721
Volume 7 (2): 271–287
Copyright© 2006
SAGE Publications
London, Thousand Oaks CA,
New Delhi

Forum Section

Methodological Pitfalls of Convergence Analysis

Katharina Holzinger
Hamburg University, Germany

The study of cross-national policy convergence has become highly popular in political science. The academic popularity of the topic significantly increased in the 1990s. There is an ever-growing body of research that investigates the occurrence and the underlying driving forces of policy convergence. Notwithstanding these efforts, we still have a limited understanding of the extent and causes of policy convergence. Both conceptual and methodological heterogeneity impose important restrictions on the comparability of the empirical findings.

Whether a study finds convergence, divergence or persistence of policies depends very much on the measurement concepts used. As an abstract statement, this seems trivial. However, Seeliger (1996) has already pointed to the fact that, in empirical convergence analyses, little attention is paid to these problems. This contribution presents a selection of the many methodological problems one is confronted with in convergence analysis. It focuses on the convergence of policies. However, similar problems may arise with respect to the convergence of institutions, culture, habits, etc. Furthermore, this contribution focuses on problems of measurement, leaving aside problems of explaining convergence. The problems will be illustrated using data from the research project 'Environmental Policy Convergence in Europe' (ENVIPOLCON).[1]

Problems of measurement of similarity

The measurement of policy convergence faces a number of problems, relating both to 'policy' and to 'convergence'. Usually, policy convergence is broadly understood as an increase in policy similarity over time. This raises two questions. First, how can the 'similarity' of policies be measured? Second, how is its 'increase over time' conceptualized? I start with problems of measuring the similarity of policies before coming back to problems related to measuring convergence.

Comparing policies

If we talk about cross-national policy convergence, we are usually referring to a certain policy field or to a single policy measure in a certain policy area. This policy (field) potentially exists in several countries, and the question is whether it has become more similar in the observed countries and time period. However, policy fields and single policies are highly complex subjects, encompassing many dimensions that can be compared in order to assess their similarity. In the literature, we generally find a broad list of policy dimensions on which convergence might occur, including, for instance, policy output or content, policy style or process, as well as policy outcomes (Bennett, 1991: 218). Another differentiation is suggested by Dolowitz and Marsh (1996: 349–50), which includes 'policy goals, structure and content; policy instruments or administrative techniques' (see also Heichel et al., 2005: 831).

With respect to methodological questions, one important distinction is between policy *output* and policy *outcome*. In the former we are interested in comparing political measures as such, whereas in the latter we are interested in comparing their effects, such as the effects of a certain policy measure on unemployment levels. Outcome data are often easily available, usually in metrical form and often in time series, and are directly comparable. Comparing output data over time and countries, however, poses more problems. Therefore, I will concentrate on policy output data here, in order to show the additional complications in comparing similarity. The methodological problems discussed later on, however, arise with both output and outcome data.

A policy field, such as industrial policy, agricultural policy, internal security policy or environmental policy, consists of a large number of policy measures. Assessing the similarity of a policy field implies a comparison of the composition of the whole field: Does the same policy field in two countries consist of a similar number and type of policy measures? Assessing the similarity of single policy measures implies, for example, asking about the existence of such a measure, the exact goals pursued with it, or the instruments used. Thus, to assess policy similarity, one can compare not only the policy repertoire of a country but also the presence of certain policies, the instruments used, and sometimes the exact metrical setting of the policy (such as a tax or a limit value). Which dimensions of a policy measure are present and important will vary across policy fields.

For the purpose of this contribution, I use data on environmental policy measures that were collected to investigate convergence effects on three different policy dimensions, namely, the *presence of a policy*, the *policy instruments* applied, and the exact *settings* of these instruments. The three dimensions vary with respect to the 'extent' of their similarity, as well as with respect to the quality of the data: the presence of a policy represents a low level of similarity and can be measured only with categorical data (yes/no); employing the same instruments represents a higher degree of similarity, also measured by categorical data (type of instrument); and the same setting of a policy implies the highest extent of similarity that can be expressed by metrical data.

To assess the similarity of policy fields, some form of aggregation of single policy measures is needed. Moreover, for the assessment of single policy measures it is necessary to aggregate the various dimensions compared. How deeply similarity will be analysed is, of course, a matter of the research question. For some purposes it is sufficient to compare the similarity of a policy field by looking only at its composition of single measures. For other purposes, such as regulatory competition research, data on the exact setting of certain parameters, such as taxes, might be needed. For still other purposes it might be necessary to aggregate several dimensions of a policy measure. The last purpose poses the problem of aggregating different kinds of data, such as categorical, ordinal and metrical, within one similarity scale. In the case of the data used for exemplification here, aggregation has taken place at the level of dimensions of policy measures, of single policies composing the environmental policy of a country, and of countries. The method used will be developed further below. So far, my goal has been simply to point to the general problems in comparing policy fields or policy measures and assessing their similarity.

Assessing the similarity of policies in the case of non-existence of a policy

A second general problem confronts any study of policy convergence: assessing the similarity of policies in the case of non-existence of a policy. At first sight this may seem to be a very specific technical problem. However, decisions about how to deal with this can dramatically change the results of convergence analysis.

What is the problem here? Assume that we want to know whether a certain policy in a number of countries has converged over a given period of time. We need to compare the policies in these countries for at least two points in time. Whereas at the second point in time (the end of the period) such a policy exists in all or most countries, at the first point in time (the start of the period) only a few countries already have such a policy and others follow only later. If we compare two countries from the set where one has the policy and the other does not, it is obvious that they should be rated as dissimilar at that point in time. However, how should we assess two countries neither of which has the policy yet? Are they similar or dissimilar?

At first sight one is tempted to say that they are similar in that they do not have a policy. However, does it make sense to talk about policy similarity in the absence of policies? Moreover, if one thinks of the metrical dimension of a tax or limit value, the problem becomes more sophisticated. If countries are rated fully similar only when their limit values are exactly equal and less similar when their limit values are different, how can one rate them as equally similar when they do not have a value at all? It would be more natural to rate them as most dissimilar – although it is definitely not possible to assign a metrical value to them.

Although philosophically there is no true answer to this problem, in practice there is a better and a worse answer. This will be shown using the environmental policy data set. This data set consists of 40 policy measures in 24 countries whose similarity is assessed at four points in time (1970, 1980, 1990, 2000). The data given in Table 1 relate to the dimension *presence of policy*. The table shows the aggregated data for the similarity of all 40 policies in all 24 countries at four points in time, measured as percentages of 'complete' similarity (that is, all 40 policies exist in all 24 countries). Additionally, it gives the data for convergence, that is, the change in similarity within each of the three decades. In version 1, pairs of countries are rated *dissimilar* when neither yet has a policy; in version 2 these country pairs are rated *similar*.

The table shows that the similarity of policies is clearly overestimated in version 2. In 1970, most countries did not yet have much of an environmental policy, thus most of them are 'similar'. In version 2, similarity is higher in 1970 than in the subsequent decades, so there is almost no convergence, and even divergence, over time. This does not at all fit the picture if one looks at the development of environmental policies in the 24 countries over the past 30 years: the countries have adopted more and more environmental policies and therefore their policy repertoire has become ever more similar. Version 1 describes the development much better: similarity of the policy repertoire increases for each point in time, and there is thus substantial convergence. One should be aware, however, that here lies a serious problem for any kind of convergence analysis, given the drastic difference in the results, which can be seen in Table 1.

Table 1 Assessment of policy similarity in the case of non-existence of policies

	Policy similarity				*Policy convergence*		
	1970	*1980*	*1990*	*2000*	*1970s*	*1980s*	*1990s*
Presence of policy – version 1	.03	.12	.30	.65	.09	.20	.51
Presence of policy – version 2	.82	.73	.67	.74	–.09	–.06	.07

Measurement of convergence

Various notions of convergence have been developed that imply different conceptions of measurement. The most basic notion is the concept of so-called σ–convergence. According to this concept, the degree of convergence increases with the extent to which the policies of different countries have become more similar over time. Thus, convergence is the decrease in standard deviation from time t_1 to t_2 and can be measured using the coefficient of variation (see Botcheva and Martin, 2001: 5; Holzinger and Knill, 2004: 30; 2005: 776; Martin and Simmons, 1998: 753f.).

In addition to σ–convergence, various concepts to assess the degree of cross-national policy convergence have been applied (for an extensive discussion, see Heichel et al., 2005: 831ff.). The concept of β–convergence measures the extent to which laggard countries catch up with leader countries over time, implying, for instance, that the former strengthen their regulatory standards more quickly and fundamentally than the latter. By contrast, γ–convergence is measured by changes in country rankings with respect to a certain policy. According to this approach, convergence increases with the extent to which country ranks change over time. Finally, δ–convergence measures the change in the distance of a given policy from a certain reference policy, e.g. the policy of the best-performing country in a set. This concept presupposes that policies can be unambiguously rated as better or worse, which is of course not always easy to assess.

The concepts of σ–, β–, γ– and δ–convergence imply different definitions and hence reference points for assessing changes in policy similarity over time. Depending on the type of convergence investigated, empirical results might be interpreted very differently. Evidence of σ–convergence, for instance, does not necessarily mean that there is also γ–convergence or δ–convergence. Evidence of β–convergence does not imply that there must also be σ–convergence: the fact that laggard countries change more fundamentally than leader countries is not a sufficient condition for a decrease in variance across all countries. It is therefore hardly surprising that we arrive at partially different assessments with regard to the degree of convergence, depending on the concrete concept applied.

Problems of σ–convergence

I shall focus on some difficulties with the most common approach, namely σ–convergence. The classic measurements are the variation coefficient for assessing changes in similarity, and the movement of the mean over time for assessing the direction of convergence whenever the policies concerned allow for such a judgement (Holzinger and Knill, 2005: 777). I start with an example of policy development in order to show the problems of the coefficient of variation and then go on to another example in order to show the problems with the development of the mean as a measure for direction. Finally, I will turn to another problem inherent in convergence analysis: saturation effects.

Figure 1 shows the policy development for industrial discharges of zinc into surface water. To illustrate the policy development, Figure 1 displays the regulations for eight countries that form a 'regulatory corridor', that is, they

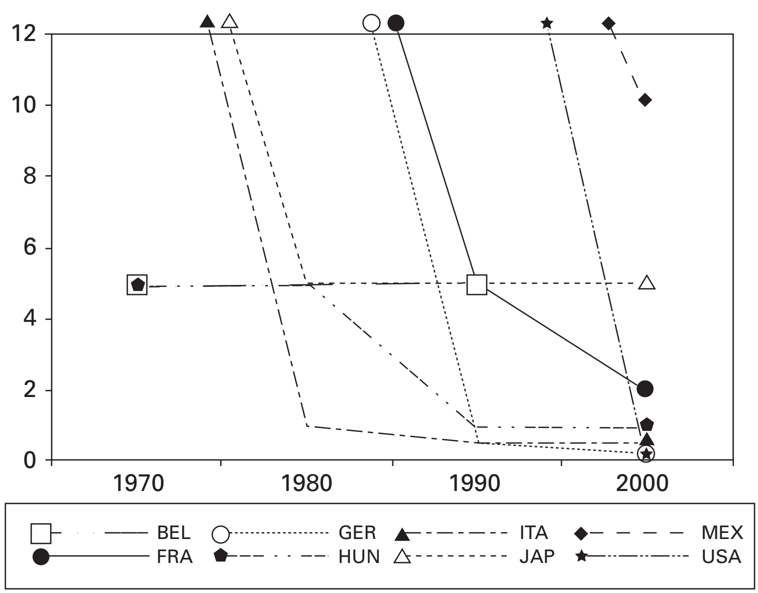

Figure 1 Limit values for zinc emissions into surface water, 1970–2000 (mg/litre).

Table 2 Limit values for zinc emissions into surface water, 1970–2000

Coefficient of variation	*1970*	*1980*	*1990*	*2000*
Perspective 1	.00	.54	.64	.18
Perspective 2		.63	.95	.95
N	2	5	10	16

comprise the maximum and minimum limit values among the whole 24-country sample for each point in time. The sample seems gradually to converge toward stricter standards in 2000. However, the coefficient of variation (Table 2, perspective 1) reveals a clear pattern of divergence throughout the whole period under study. This result is strikingly different from the optical impression.

An important reason for divergence as expressed by the coefficient of variation is to be found in the number of new policy adoptions during the 1980s and 1990s. As a result, the range of values and their variance changed significantly over time. But, even if only those countries for which a value already existed in the previous time period (Table 2, perspective 2) are taken into account, no clear tendency of σ–convergence can be observed.

A second reason for the difference between the impression from the graphical depiction of limit values and the coefficient of variation is the latter's sensitivity to outliers. The picture of divergence in the national regulation of industrial discharges of zinc is modified if we analyse the variation by an alternative instrument: box-plots. Box-plots display the variation of all values in one box for each point in time. Each box includes 50% of all values; the upper and the lower quartiles delimit the box on both sides. Outliers and extreme values are displayed separately. This way, they do not influence the length of the box. The line across the middle of each box is the median. The length of the box represents the degree of variation: the longer the box, the less similar the values on this variable. Decreases in the length of a box can be interpreted as an occurrence of σ–convergence.

Figure 2 applies this instrument to the case of zinc regulation. Whereas, for the first periods under investigation, both the box-plot analysis and the coefficient of variation indicate divergence, the interpretation offered by both instruments points in different directions for the last period. The left-hand side of Figure 2 indicates a clear trend of policy convergence during the 1990s for the core group of countries (the box is much shorter in 2000). In 2000 there is an obvious outlier (Mexico), which is responsible for the impression of divergence when the coefficient of variation is used as an indicator. The

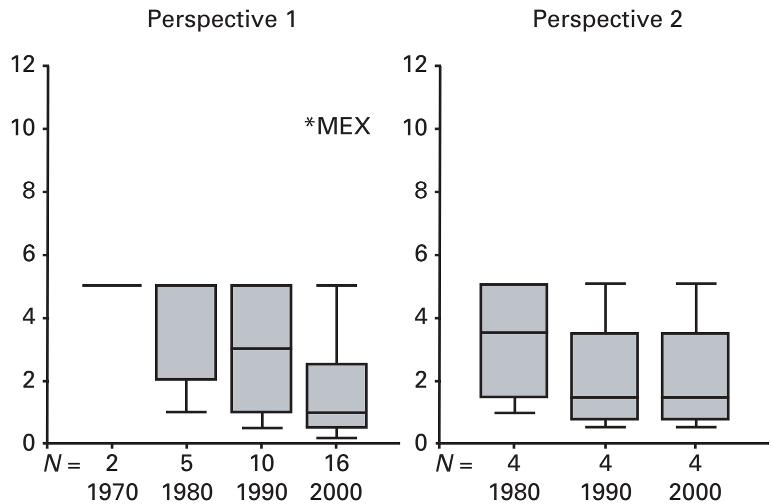

Figure 2 Industrial zinc discharges in surface water, 1970–2000 (mg/litre).

picture of different interpretations is confirmed, irrespective of the selected convergence perspective. Whereas the left-hand side of the figure includes all available values for each point in time with the number of countries changing over time (perspective 1), the right-hand side of the figure is based on perspective 2, taking into account only those four countries that had already established a standard in 1980 (the value for Belgium is missing in 2000). The figure shows that these early adopting countries had already converged during the 1980s.

The next example shows that the development of the mean can be very misleading as well. The development of the mean limit value for coliform bacteria in bathing water is as follows: 1000 per 100 ml (1970), 5500 (1980), 7167 (1990), and 8000 (2000). The mean clearly indicates that limit values have become less strict over time: the more coliforms permitted, the laxer the standard.

This is again in sharp contrast to the impression one gets from looking at the graphical display of the developments for all countries. Figure 3 is a bar chart of the limit values for all countries. It exhibits a clear convergence towards a standard of 10,000 coliforms per 100 ml in 2000 (the black bars). Some countries had already introduced this standard earlier (the UK, France, Greece, Portugal, Spain); others introduced it only in the decade before 2000 (Austria, Finland, Ireland, Netherlands, Poland). Some now use a stricter standard of 1000 or 2000 coliforms per 100 ml (Denmark, Hungary, Italy, Japan, Slovakia). In contrast to the expectation raised by the development of

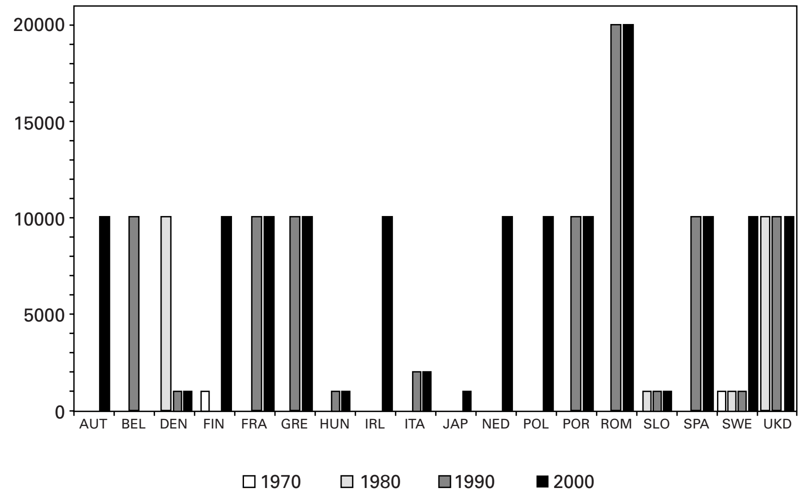

Figure 3 Development of limit values for coliforms in bathing water, 1970–2000 (per 100 ml).

the mean, only two countries lowered their standards, namely Finland and Sweden (from 1000 to 10,000). Together with the introduction of a very lax standard of 20,000 by Romania during the 1980s, this obviously accounts for the increase in the mean in 1990 and 2000. In 1970, only two countries had a standard – the strict value of 1000 in Finland and Sweden. The increase in the mean in 1980 follows from the introduction of the 10,000 value in Denmark and the UK. Thus, the overall increase in the mean is the result of wide variation in this limit value (which is a consequence of scientific uncertainty in this matter). Nevertheless, there is clear convergence to the top: in 1970, only two countries had any standard, whereas, in 2000, 11 countries applied the 10,000 standard, 5 applied a stricter one and only 1 country applied a laxer one. The development of the mean gives an inadequate impression of the direction of convergence.

The examples show that both the coefficient of variation and the mean as standard measurement concepts for σ–convergence have a number of problems. First, both are sensitive to outliers and extreme values and can thus give a false impression of what actually happened. Second, they cannot account very well for 'newcomer' countries, that is, countries that do not have a policy from the start. Newcomers may increase the variation in limit values, despite the fact that early adopters are converging and that more countries are adopting a policy, which can itself be interpreted as convergence. Third, and this has not been made explicit so far, the coefficient of variation and the mean are measures that can be used only with metrical data. Other dimensions of policy convergence, such as the adoption of a policy in ever more countries or the growing similarity of the instrument used, cannot be grasped with these measures. Thus, the aggregation of several dimensions of a policy in order to have a more adequate measurement of convergence is impossible with these concepts too.

A new measurement concept for σ–convergence

Within the ENVIPOLCON project, a more refined measurement concept of σ–convergence has been developed that can overcome these weaknesses of the classic approach (see Holzinger et al., 2006). Since the figures presented in Table 1 and in later examples are based on this new approach to the measurement of similarity and σ–convergence, it will be briefly introduced here.

The 'pair approach' is based on the comparison of policy changes across a number of countries. If the direction of convergence is not of interest, the most direct approach to convergence measurement is to compare each country individually with each other country in the sample. The pairwise comparison is the basic starting point for the study of σ–convergence. The units of analysis are country pairs and not single countries. Consequently, the concept of convergence implies an increase in policy similarity between a certain pair of countries over time. The use of country pairs or dyads is new to the study of policy convergence, although it is common in other research areas, such as the study of international conflict (Kinsella and Russett, 2002; King, 2001).

The pair approach involves several advantages for the study of σ–convergence. First, the pair approach comprises any shift of convergence or divergence between all pairs of countries. This way, sensitivity to outliers and newcomers is avoided. Second, it can be used for both categorical and metrical data. Thus, using the pair approach, the various dimensions of policies can be integrated into one measure. Third, since it is not based on aggregate figures, such as the coefficient of variation, it allows for the use of a convergence variable as the explanandum in a quantitative model.

How is pairwise similarity calculated? If a policy is present for both countries, they get a score of 1 (similarity), otherwise a score of 0 (dissimilarity). The same applies if both countries use the same instrument. For the setting dimension, a normalized metrical score from 0 to 1 is applied, based on differences between the limit values of country A and country B, leading to a similarity scale between 1 (limit values are identical) and 0 (country pair with the most dissimilar setting values). For all other values, gradual similarity is assessed by weighting the distance between two settings with the maximum distance for each item and for each point in time. These scores can subsequently be aggregated over the dimensions of a single policy, over all policies and over all countries of the sample. This way we get the overall similarity of the policy field at one point in time. The score can then be expressed as a percentage of the maximum score for complete similarity of all policies in all countries. Finally, convergence is expressed as the change in similarity within a decade. These measures for similarity and convergence are given in Table 1 and in further tables below.

The saturation effect

There is one more problem that is inherent in any analysis of convergence: the saturation effect. If convergence of policies takes place in a certain group of countries and a given period of time, this process has its natural endpoint when all countries are fully similar with respect to all policies (or a single policy). During the process, if a number of countries from the sample already have fully similar policies, they cannot converge any further. That is, the potential for further convergence within the whole group shrinks. The more similar policies already are, the more the potential for further convergence decreases within the respective group. In short, over time, saturation takes place.

Since policies and countries that are already fully similar cannot converge any further, convergence of the other countries is underestimated if we do not take into account the saturation effect. To avoid this, policies that have become completely similar for a country pair A and B in an earlier period and remain stable in the subsequent period should be excluded from the policy sample in the calculation of subsequent convergence. That is, the aggregated convergence score for this pair is weighted not by the number of all policies but by the number of policies that are not yet fully similar.

The environmental data set can be used to show the difference in the results when the saturation effect is controlled for. It is to be expected that the saturation effect is greater in later periods – given that there is a general convergence trend. Policy convergence for the *presence of policy* dimension is .09 for the 1970s, .20 for the 1980s and .51 for the 1990s if the saturation effect is controlled for (see Table 1, version 1). If the saturation effect is not controlled for, the respective figures are .09, .17 and .35. Thus, there is in fact not much difference in the 1970s – the convergence figures are almost the same. In the 1990s, however, there is a clear difference: convergence rises from 0.35% to 0.51%.

Selection problems

In this section I turn to the problems of sample selection. This is, of course, a general problem not specific to convergence analysis. In fact, case selection is more important in qualitative research because the number of cases is often small (King et al., 1994: 115ff.). However, selection problems also arise in quantitative studies, and thus also in quantitative convergence analysis. Seeliger (1996: 289ff.) has pointed to the fact that in empirical convergence studies the problem of sample selection is often neglected.

In the analysis of cross-national policy convergence, several kinds of samples are of interest: the selection of policy fields, the selection of the single policies that comprise a policy field, the selection of countries, and the selection of time periods. Policy fields will be chosen according to research interest, and selection becomes important whenever several fields are to be compared. Similarly, the selection of countries may correspond to a region that is the focus of the project, or it may correspond to a sample that represents the world or a region. Single policies comprising a policy area are usually not a 'census' but a sample, and the same is true for time periods or points in time chosen for measurement. Especially with respect to time periods, data availability often plays a role in the selection process. It is rarely possible for a given policy field and a given group of countries to obtain data for the full time period over which the policy exists in at least one of the countries. Moreover, random selection in order to achieve representative samples will often be impossible too. This is especially true for policies for which no 'full list' exists. I will present some examples showing how sensitive convergence analysis can be to even minor changes in the sample.

Country sample

Table 3 gives the coefficient of variation of three environmental policy measures for two points in time, 1990 and 2000. The change in the coefficient indicates convergence or divergence of these policies among a group of countries that have such a policy. The three policies are the sulphur content of gas oil and NO_x (nitrogen oxides) and HC (hydrocarbon) emissions from passenger cars. The table shows that taking away only one or two countries from the samples of 17 and 14, respectively, may affect the direction of policy change. For the sulphur content of gas oil, there seems to have been clear convergence of policies if Bulgaria is removed from the sample. Including Bulgaria changes the picture completely, because Bulgaria is an outlier in 2000. For the emissions from passenger cars, the full sample seems to have converged between 1990 and 2000. Removing Hungary and Ireland results in divergence for the remainder of the countries. Hungary and Ireland were outliers in 1990 and caught up during the 1990s. Thus, small changes in the group can have large effects on the results of convergence analysis.

Table 3 Effects of country selection

Example policy	*1990*	*2000*	*Convergence or divergence?*
Sulphur content of gas oil			
– 17 countries	.77	.98	Divergence
– without Bulgaria	.49	.18	Convergence
NO_x emissions from passenger cars			
– 17 countries	.92	.38	Convergence
– without Hungary and Ireland	.20	.39	Divergence
HC emissions from passenger cars			
– 14 countries	.72	.53	Convergence
– without Hungary and Ireland	.11	.55	Divergence

Time period

The time frames of convergence studies are of 'paramount importance' (Seelinger, 1996: 296) for the interpretation of their results. For the environmental policy convergence project, a time period of 30 years was chosen, which is longer then usual in convergence analysis. In this case the time period covers almost the full span of the history of environmental policy, which began in most countries in the late 1960s; it comes close to a census. Table 4 shows the variation in the extent of convergence if each of the three decades, two 20-year periods or the full period of 30 years are observed.

The table indicates that convergence was lowest during the 1970s and highest during the 1990s, with the 1980s in-between. Accordingly, the first 20-year period shows less convergence than the second one. The whole 30-year period shows less convergence than the 1990s and also less than the second 20-year period (1980s and 1990s). Thus, convergence clearly increased over the three decades. The selection of a different time period, e.g. only the period between 1980 and 2000 or only the 1990s, would obviously lead to a different impression of the amount of environmental policy convergence than would the full time span.

In the above example, the time period is represented by four points in time. This allows for four cross-sectional analyses of similarity or several cross-sectional analyses of convergence over various periods (see Table 4). Although yearly time series data might seem more attractive, their use can be problematic in convergence analysis. This depends very much on the object of convergence analysis or, more precisely, the dependent and independent

Table 4 Selection of time period

	All policies	Presence of policy	Instruments	Settings
1970s	.10	.09	.16	.06
1980s	.18	.20	.21	.10
1990s	.37	.51	.36	.18
1970s and 1980s	.14	.14	.19	.08
1980s and 1990s	.28	.36	.29	.14
1970s to 1990s	.22	.27	.24	.12

variables. For some variables – for example when economic growth is the dependent variable and investment the independent variable – yearly data may be easily available and there may be sufficient variation over time in order to apply time series analysis sensibly. When the dependent variable is 'policy output', however, data will usually have to be collected first, which the available resources do not always permit. Moreover, policies do not change on a yearly basis and therefore there is not enough variance over time for a yearly time series. A time span of 5 or 10 years is much more appropriate here. If the time period observed is not very long, time series analysis is impossible.

The selection of policy measures

The 40 policy measures chosen for the assessment of environmental policy convergence are a non-representative and non-random sample. Drawing a random sample is impossible because environmental policy measures are an open list ('the universe of cases is not clearly specified', King et al., 1994: 125). The selection of the sample was guided by theoretical interest. The main explanatory factors to be tested were the effect of international institutions and the effect of regulatory competition on the extent of environmental policy convergence. For this reason an equal number of policies were selected for which an obligatory international policy exists and for which such an international policy does not exist, as well as an equal number of policies that are trade related and that are not trade related. It was expected that stronger convergence would occur for obligatory international policies and for trade-related policies. It was also expected that convergence would be most pronounced with respect to the dimension of *presence of policy*, and least pronounced with respect to the *setting* dimension.

Table 5 shows that convergence varies greatly over the various subgroups of policies. It is highest for the *presence of policy* dimension of policies for which

Table 5 Selection of policy measures

Dimension	*Obligatory international policy exists*	*No obligatory international policy exists*	*Policy is trade related*	*Policy is not trade related*
Presence of policy	.44	.20	.32	.19
Instruments	.40	.13	.31	.08
Settings	.18	.07	.14	.05

international harmonization exists (.44) and it is lowest for the *settings* dimension of policies that are not trade related (.05). The pattern fully 'converges' with the theoretical expectations: obligatory policies converge more than non-obligatory policies; trade-related policies converge more than non-trade-related policies; and the *presence of policy* dimension converges more than the *instruments* and the *settings* dimensions. This is, however, not the point to be made here. The pattern of subgroups makes clear that the extent of convergence varies greatly across policy measures. The researcher has to be very cautious: each arbitrary (and non-random) selection of policies might yield very different results and any generalizations are problematic.

In general, problems of selection will be severe in convergence analysis. In many instances, we will have only small numbers of cases (countries or policies) or short time periods, there will be problems of data availability, time series data will not be available or cannot be used in a sensible way, and random selection will be impossible. The above examples have shown how great the effects of minor changes in the samples can be. This makes it very important to reflect carefully on the selection of time frames, countries and policies in designing convergence research and to keep these problems in mind in interpreting its results.

Conclusion

Because policy convergence is a growing field of research in political science, it is important to point to a number of methodological problems arising in the measurement of convergence. The measurement of convergence includes as a first step the measurement of similarity and as a second step the measurement of convergence. With respect to the first step, I have dealt with the difficulties of comparing policies and of measuring and aggregating various dimensions of policies. I also discussed a new method to deal with these problems: the pair approach. With respect to the second step, I discussed four conceptions of convergence. I demonstrated some problems with the most common conception, σ-convergence, related to the use of the coefficient of variation and the mean. The pair approach can avoid these problematic effects.

Some of the methodical problems discussed here, such as the problems of selection, are of a general nature. Other problems, however, are specific to policy convergence analysis. First, there is the problem of assessing the similarity or dissimilarity of policies in situations where some countries do not yet have a policy. I have shown that the decision to rate them as similar has severe effects on the result: it can even eliminate convergence in highly converging policy fields. Second, there is some kind of saturation effect. The more policies have already converged, the less potential remains for further convergence. If this is not taken into account, convergence in later periods of time will be underestimated.

Taken together, there are many pitfalls when measuring convergence. These problems may lead the researcher to overestimate or underestimate convergence effects. It is even possible to see divergence with one kind of measurement and convergence using another, or to see convergence to the top or bottom where another measure produces the opposite result. The use of appropriate conceptions of convergence and appropriate measures and indicators of similarity and convergence, as well as the careful selection of time periods, and country and policy samples, is of paramount importance for future convergence research. To make results comparable, these methodological questions have to be discussed more explicitly among convergence researchers.

Notes

Funding of this research by the European Commission within the fifth framework programme on socioeconomic research is gratefully acknowledged.

1 The ENVIPOLCON research team includes the universities of Konstanz, Nijmegen, Hamburg, Berlin (Free University) and Salzburg.

References

Bennett, Colin (1991) 'What Is Policy Convergence and What Causes It?', *British Journal of Political Science* 21(2): 215–33.

Botcheva, Liliana and Lisa L. Martin (2001) 'Institutional Effects on State Behaviour. Convergence and Divergence', *International Studies Quarterly* 45(1): 1–26.

Dolowitz, David P. and David Marsh (1996) 'Who Learns What from Whom? A Review of the Policy Transfer Literature', *Political Studies* 44(3): 343–57.
Heichel, Stephan, Jessica Pape and Thomas Sommerer (2005) 'Is There Convergence in Convergence Research? An Overview of Empirical Studies on Policy Convergence', *Journal of European Public Policy* 12(5): 17–840.
Holzinger, Katharina and Christoph Knill (2004) 'Regulatory Competition and Regulatory Cooperation in Environmental Policy: Individual and Interaction Effects', *Journal of Public Policy* 24(3): 25–47.
Holzinger, Katharina and Christoph Knill (2005) 'Causes and Conditions of Cross-national Policy Convergence', *Journal of European Public Policy* 12(5): 75–796.
Holzinger, Katharina, Christoph Knill and Thomas Sommerer (2006) 'The Impact of International Institutions and Trade on Environmental Policy Convergence in Europe', unpublished manuscript.
King, Gary (2001) 'Proper Nouns and Methodological Propriety: Pooling Dyads in International Relations Data', *International Organization* 55(2): 497–507.
King, Gary, Robert Keohane and Sidney Verba (1994) *Designing Social Inquiry. Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
Kinsella, David and Bruce Russett (2002) 'Conflict Emergence and Escalation in Interactive International Dyads', *Journal of Politics* 64(5): 1045–68.
Martin, Lisa L. and Beth A. Simmons (1998) 'Theories and Empirical Studies of International Institutions', *International Organization* 52(4): 729–57.
Seeliger, Robert (1996) 'Conceptualizing and Researching Policy Convergence', *Policy Studies Journal* 4(3): 287–306.

About the author

Katharina Holzinger is Professor of Government at Hamburg University, Institute of Political Science, Department of Social Sciences, Allende-Platz 1, D-20146 Hamburg, Germany.
Fax: +49 40 42838 7661
E-mail: holzinger@sozialwiss.uni-hamburg.de